ECE 600 MA519

Application: Allocation of capital, risk

One of Bouman's favorite topics

Random Variables

Real things are not well-modeled by deterministic processes

- Let X be a random variable on $\mathbb{R}$, then
  - X is usually denoted by an upper case letter.
  - The cummulative distribution function (always exists) is given by

$$P\{X \leq x\} = F_X(x)$$

  - If the probability density function (doesn't always exist) exists, it is given by

$$p_X(x) = \frac{dF_X(x)}{dx}$$ – Absolutely Continuous

so that

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= F_X(x_2) - F_X(x_1) \\ &= \int_{x_1}^{x_2} p_X(\tau)d\tau \end{aligned}$$

  - The expectation of X is given by

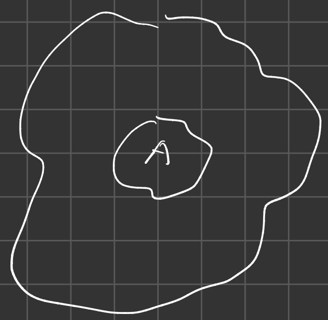
$$E[X] = \int_{-\infty}^{\infty} \tau p_X(\tau)d\tau$$

or more precisely by the Riemann-Stieltjes integral

$$E[X] = \int_{-\infty}^{\infty} \tau dF_X(\tau)$$

if it exists.

MA 504
MA 538



Barrells

$A \in \mathcal{B}$ $\qquad$ $A \subset \Omega$

$\mathcal{B} \neq 2^{\Omega}$

$A \subset \Omega$

$0 \le P(A) \le 1$
$P(\Omega) = 1$
$P(\emptyset) = 0$
$P(A^c) = 1 - P(A)$

$A \cap B = \emptyset$
$P(A \cup B) = P(A) + P(B)$

$w \in [0, 1]$
$P(\{w\}) = 0$
$\bigcup \{w\} = [0, 1]$

$1 = P(\bigcup_w \{w\}) \neq \sum_w P\{w\} = \sum_w 0$

definition of disjoint?

$\mathcal{B}$

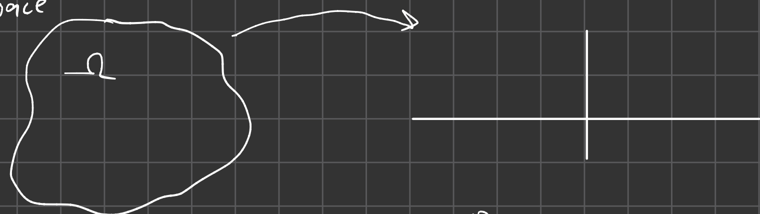
$(\Omega, \mathcal{B}, P) \rightarrow$ defines a probability space

$P : \mathcal{B} \rightarrow [0, 1]$
$0 \le P(A) \le 1$

Basic Properties of Mathematics

event space

X - maps from Ω to real line



$(\Omega, \mathcal{B}, P)$

$X: \Omega \to \mathbb{R}$
$X(\omega) \quad \omega \in \Omega$

- Ω is what can happen
- random variable depends on the event

CDF: - cumulative distribution function

$\{\omega \in \Omega : X(\omega) \leq \lambda\}$
$P(A_\lambda)$

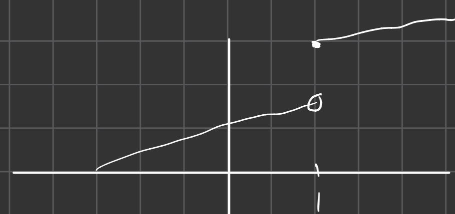
$\triangleq P\{X \leq \lambda\} = F_X(\lambda)$

R.V. always function of element from the probability space

Marco Question: should be X^{-1} ???

CDF Properties
1) monotone non-decreasing
2) $\lim_{\lambda \to \infty} F_X(\lambda) = 1$
$\lim_{\lambda \to \infty} F_X(\lambda) = 0$
3) right-hand continuous
countable infinitely many discontinuities

CDF doesn't have to be continuous:



$\to$ right hand continuous

Diagonalization Proof

$$\frac{dF_X(\lambda)}{d\lambda} = p_X(\lambda)$$

$$F_X(\lambda) = \int_{-\infty}^{\lambda} p_X(\sigma)\, d\sigma$$ - not always true:

$$p_X(\lambda) = \frac{\partial}{\partial x} \int_{-\infty}^{\lambda} p_X(\sigma)\, d\sigma$$ — always true

Cantor Functali ??

Deterministic versus Random

- Let X and Z be random variables, and let $f(\cdot)$ be a function from $\mathbb{R}$ to $\mathbb{R}$
  - Is $Y = \underbrace{f(X)}$ a random variable? Yes

    function of a random variable
  - Is $\mu = E[X]$ a random variable? No - deterministic
  - Is $\hat{X} = E[X|Z]$ a random variable? Yes

    $= f(Z)$

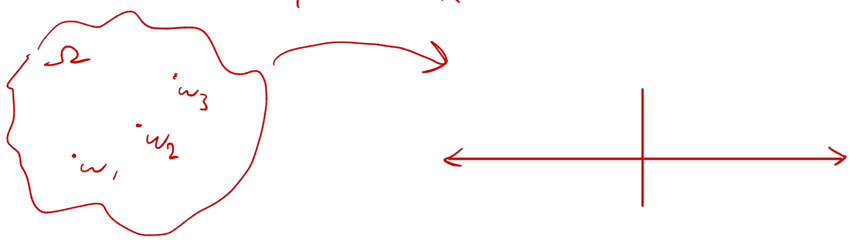
    conditional expectation forms the foundation of estimation

$E[Z|Z] = Z$

$E[f(Z)|Z] = f(Z)$

Random Variable : function of $w \rightarrow Y(w)$

$Y : \Omega \rightarrow \mathbb{R}$



Uppercase: function of w ✱

Properties of Expectation

Know for future exams ✱

- Expectation is linear

  Expectation defined by an integral, therefore it is linear

$$E[X+Y] = E[X] + E[Y]$$

- What is $E[E[X|Y]]$ equal to?

  since we are considering all y

$$E[E[X|Y]] = E[X]$$

  very useful in practice/research

- What is $E[X|X,Y]$ equal to?

  and

$$E[X|X,Y] = X$$

  useful when you have a lot of conditions

- When X, Y, and Z are (jointly) Gaussian

  recall equation for Gaussian distribution

$$f(Y,Z) = E[X|Y,Z] = aY + bZ + c$$

  zero when zero mean

  linear function - affine

  for some scalar values a, b, and c.

  review multivariable cdf's & pdf's

$E[X|Y] = f(Y)$

$f(y) = \int x\, p(x|y)\, dx \rightarrow E[X|Y] = f(Y)$

some Y

unless otherwise specified, X is probably a countably infinite set

$p(x,y) = p(x|y)\, p(y)$

2-D Discrete Space Random Processes

↳ collection of random variables

- Notation
  - X_s is a pixel at position $s = (s_1, s_2) \in \mathcal{Z}^2$ - discrete set
    vector index
  - S denotes the set of 2-D Lattice points where $S \subset \mathcal{Z}^2$
- Definitions
  - Mean $\mu_s = E[X_s]$ - may not exist depending on pdf
  - Autocorrelation $R_{sr} = E[X_s X_r]$ if i.i.d. Gaussian & $s \neq r$ → independent
  - Autocovariance $C_{sr} = E[(X_s - \mu_s)(X_r - \mu_r)]$

    if 0 mean, same
  - A process is said to be **second order** if $E[X_s]$ and $E[X_s X_r]$ exist for all $s \in S$ and $r \in S$. review "moments" in prob/stats
  - A second order random process is said to be **wide sense stationary** if for all $s \in \mathcal{Z}^2$

review what moments are

SSS is also WSS if it's second order

$$\mu_s = \mu_{(0,0)}$$

mean not function of position (mean space invariant)

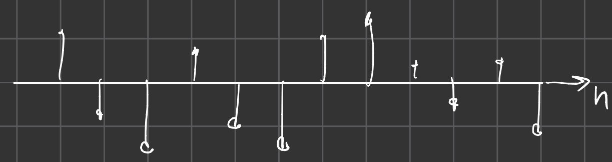
$$C_{r,r+s} = C_{(0,0),s}$$

autocovariance only function of difference vs absolute location

stationary types:
1) wide sense - needs to be second order (considered weaker)
2) strict sense - all distributional properties not a function of time

X_n $\quad n = \ldots -3, -2, -1, 0, 1, 2, 3 \ldots$

$n \in \mathbb{Z}$ - counting integers



$X_n(w)$

stationary - distribution is not a function of time (time/space invariant)

- can't tell where you are in time by looking at a particular window of the sample path, $X(w)$

- distribution fundamentally describes the random process
- for jointly Gaussian → mean and covariance determine everything about the random process

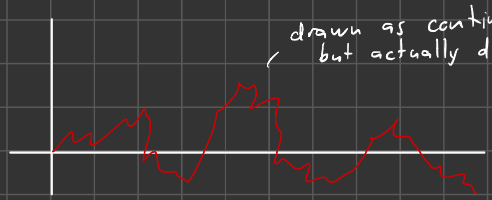
μ not function of time → μ constant

cov. not function of time → cov. only function of difference in time

Ex: $S_n = \sum_{i=1}^{n} w_i \;;\; w_i$ i.i.d. $\sim N(0,1)$

variance & standard deviation = 1

- variance & standard deviation related



$E[S_n] = 0$

$E[S_n^2] = n$ $\quad$ different index, different sum

$$E\left[\left(\sum_{i=1}^{n} w_i\right)\left(\sum_{j=1}^{n} w_j\right)\right]$$

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[w_i w_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \delta(i-j) = n$$

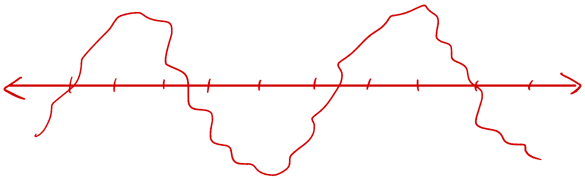
$$E[w_i w_j] = \begin{cases} E[w_i]E[w_j] = 0, & i \neq j \\ E[w_i^2] = 1, & i = j \end{cases} = \delta(i-j)$$

because variance is 1

- this process is NOT stationary

- Martingale: $E[S_n | S_m] = S_m$

A priori



stationary: can't tell where you are in time by looking at a given window
↳ statistical properties don't vary with time

2-D Power Spectral Density

Let X_s be a zero mean wide sense stationary random process. - if not stationary, the power spectral density doesn't exist

Define

$$\hat{X}_N(e^{j\mu}, e^{j\nu}) = \sum_{m=-N}^{N} \sum_{n=-N}^{N} X_{(m,n)} e^{-j(m\mu+n\nu)}$$

random variable

defines the window

- Then the power spectrum (i.e. energy spectrum per unit sample) is (which is itself random):

$$\frac{1}{(2N+1)^2} \left| \hat{X}_N(e^{j\mu}, e^{j\nu}) \right|^2$$

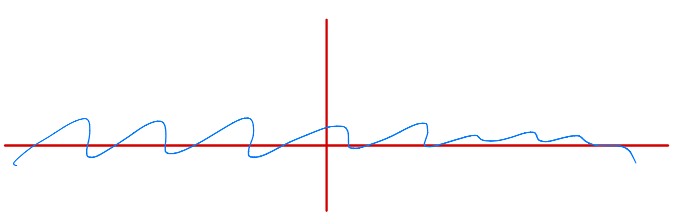
power/energy

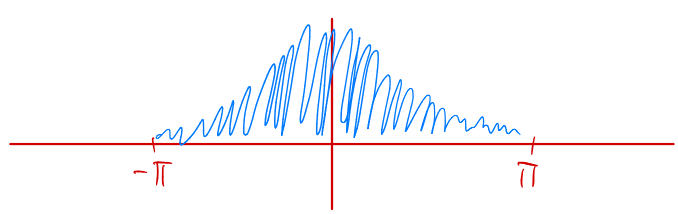
per unit time, normalizes

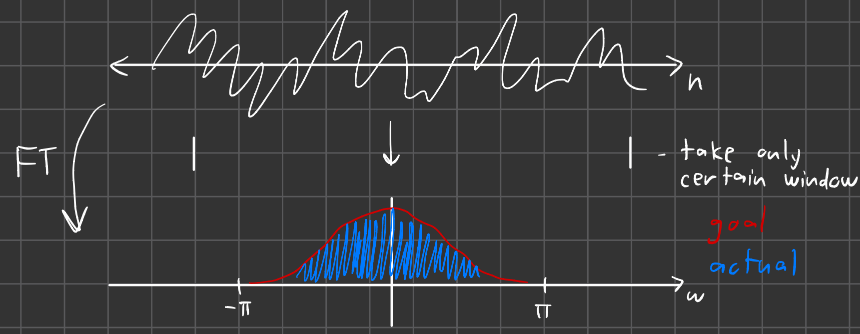
The following limit does not converge!!

$$\lim_{N\to\infty} \frac{1}{(2N+1)^2} \left| \hat{X}_N(e^{j\mu}, e^{j\nu}) \right|^2$$

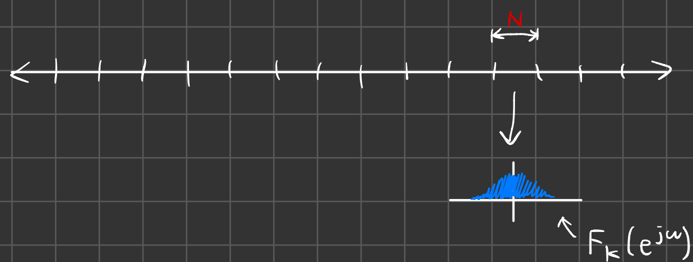
Intuition - The spectral estimate remains noisy as the window size increases.







- FT is still random



- chop signal up into pieces

- take FT of each piece

$$\frac{1}{k}\sum_{k=0}^{k} F_k(e^{j\omega}) \sim E\left[F_k(e^{j\omega})\right]$$

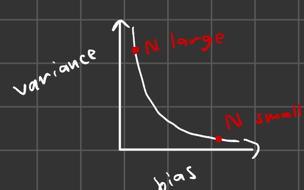
Competing Goals

1) make window big → high resolution in frequency
2) average a lot of FT's → reduce the noise

} have to sacrifice/prioritize one

noise: variance

smoothing of peaks: bias



$p(x) = \frac{1}{2} \frac{1}{1+x^2}$

$X_n \sim p(x)$ i.i.d.

not second order → expectation does not exist

Definition of Power Spectral Density

- Definition of **Power Spectral Density**

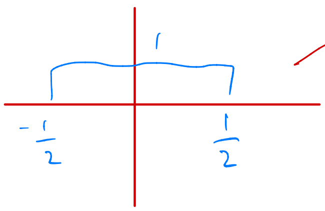
$$S_x(e^{j\mu}, e^{j\nu}) \triangleq \underbrace{\lim_{N\to\infty} \frac{1}{(2N+1)^2}}_{\text{reduces bias, increases resolution}} \overbrace{E}^{\text{reduces variance}} \left[\left| \hat{X}_N(e^{j\mu}, e^{j\nu}) \right|^2 \right]$$

Expectation removes the noise.

Lab 2 Notes:

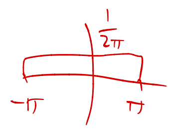
$y(m,n) = 3x(m,n) + 0.99y(m-1,n) + 0.99y(m,n-1) - 0.9801\,y(m-1,n-1)$

$R(m,n) = E[Y_{(0,0)}Y_{(m,n)}] = \sum_{m=0}^{512}\sum_{n=0}^{512} y(0,0)y(m,n) = y(0,0)\sum_{m=0}^{512}\sum_{n=0}^{512} y(m,n)$



$\rightarrow E[X_{(m,n)}] = 0$

area under any PDF = 1?



Weiner-Khintchine Theorem

- For a wide sense stationary random process, the power spectral density equals the Fourier transform of the autocorrelation

$$S_x(e^{j\mu}, e^{j\nu}) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} R(m,n)e^{-j(m\mu+n\nu)}$$

STATS19
Dr. Yip

where

$$R(m,n) = E[X_{(0,0)}X_{(m,n)}]$$

512×512 random variables

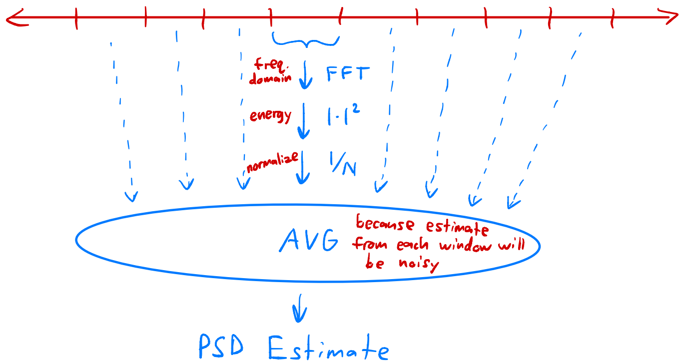
$E[AB] = E[A]E[B]$ if i.i.d, independent - both take on uniform distribution

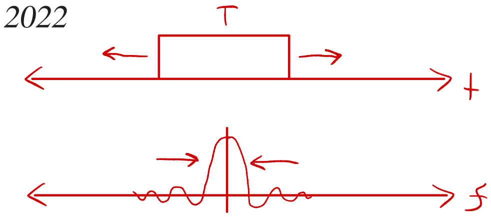
if not i.i.d, may need to use definition of expectation

isolate & solve for what you need ☆

$E[X_{(0,0)}^2] =$

could also use integral & definition





convolution with a sinc: low pass filter (smoothing)

1) Window bigger → sinc function smaller → frequencies get less smooth → more resolution → fewer blocks to average → more noise

2) Window smaller → sinc function larger → frequencies get more smooth → less resolution → more blocks to average → less noise

Estimating Power Spectral Density

- For simplicity, consider 1-D case

$$S_x(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{N} E\left[\left|\hat{X}_N(e^{j\omega})\right|^2\right]$$

How do we compute the required expectation?

- Answer: The law of large numbers (averaging)

$$E[Z] = \lim_{N\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} Z_k$$

where Z_k are independent and identically distributed (i.i.d.).

Computing Power Spectrum from Block Average

- Let X_n be signal with $0 \leq n < NK$.
- Break X_n into K parts, each of length N.

$$Y_n^{(k)} = X_{kN+n}$$

where $0 \leq k < K$ and $0 \leq n < N$.

- Compute DTFT of $Y_n^{(k)}$

$$\hat{Y}^{(k)}(e^{j\omega}) = \sum_{n=0}^{N-1} Y_n^{(k)} e^{j\omega n}$$

- Average power spectrum estimates for each block

$$\begin{aligned} S_x(e^{j\omega}) &= \frac{1}{N} E\left[\left| \hat{Y}^{(k)}(e^{j\omega}) \right|^2 \right] \\ &\cong \frac{1}{N} \left[\frac{1}{K} \sum_{k=0}^{K-1} \left| \hat{Y}^{(k)}(e^{j\omega}) \right|^2 \right] \end{aligned}$$

- So we have that

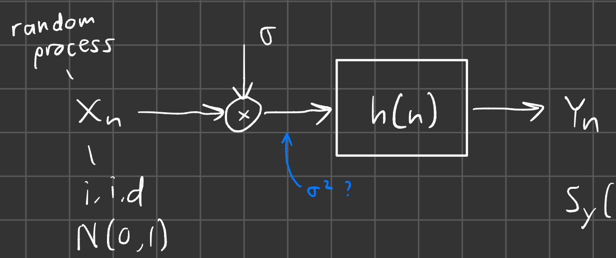
$$S_x(e^{j\omega}) \cong \frac{1}{NK} \sum_{k=0}^{K-1} \left| \hat{Y}^{(k)}(e^{j\omega}) \right|^2$$

Computing the Power Spectrum of a Random Process

$E[X_n] = 0$

$R(k) = E[X_n X_{n+k}]$ — Law of Large Numbers, Approximated with real data

$S_x(e^{j\omega}) \overset{DTFT}{\Longleftrightarrow} R(k) \rightarrow$ windowing $R(k)$ is to filter $R(s)$



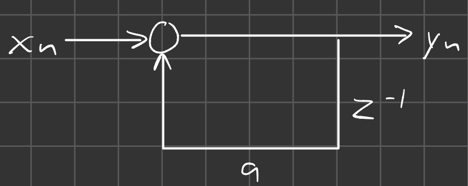
$S_y(e^{j\omega}) = ??$

$$R(k) = E[X_n X_{n+k}] = \begin{cases} 0, & k \neq 0 \\ 1, & k = 0 \end{cases} = \delta(k)$$

$\therefore S_x(e^{j\omega}) = 1 \rightarrow$ energy is uniformly distributed across all frequencies (white noise)

so, X_n is white noise

$$S_y(e^{j\omega}) = S_x(e^{j\omega}) \cdot \sigma^2 \cdot |H(e^{j\omega})|^2$$



$$Y(z) = X(z) + az^{-1}Y(z)$$

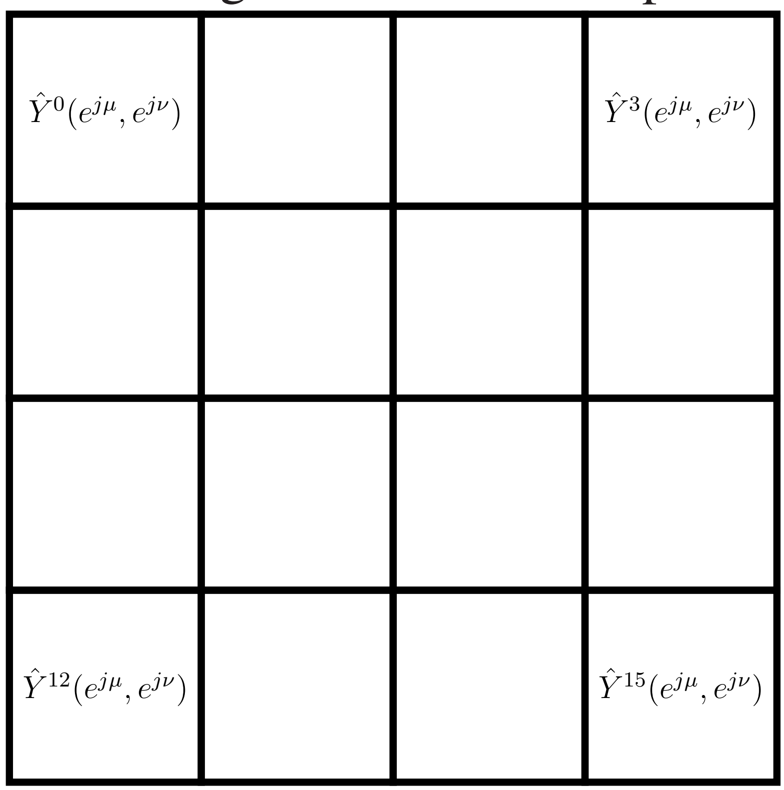
$$\frac{Y(z)}{X(z)} = \frac{1}{1 - az^{-1}}$$

$$|H(e^{j\omega})| = \left| \frac{1}{1 - ae^{-j\omega}} \right|$$

could also take FT^{-1} to compute autocorrelation of y also

Block Averaging in 2-D

- Break image into K regions each with N pixels



- For each block, compute $\hat{Y}^{(k)}(e^{j\mu}, e^{j\nu})$
- Average blocks to form power spectrum estimate

$$S_x(e^{j\mu}, e^{j\nu}) \cong \frac{1}{NK} \sum_{k=0}^{K-1} \left| \hat{Y}^{(k)}(e^{j\mu}, e^{j\nu}) \right|^2$$